



Getting Started with the CONTENTdm Flex Loader

This guide provides instructions for installing and using the CONTENTdm Flex Loader. The Flex Loader supports the import of data into CONTENTdm from the METS/ALTO format. Currently supported METS/ALTO formats are newspapers and monographs (eBooks).

Requirements

Check that you have the following before installing the Flex Loader:

- Recommended: Windows XP with Service Pack 3 or later (32- or 64-bit), or Windows Vista Business Edition with Service Pack 1 or later (32- or 64-bit).

(The following have also been tested and will work with Flex Loader: Windows XP Professional Edition, Windows Vista Ultimate, Windows Vista Enterprise, Windows 2003 R2 Enterprise, Windows 2003 R2 Standard, Windows 2008 Enterprise, Windows 2008 Data Center, Windows 2008 R2 Enterprise, Windows 7 Ultimate 32-bit.)

- Connectivity to <http://www.worldcat.org>. (Click the link or open a browser and type in the URL to check.)
- CONTENTdm Server version 5.1 or later
- Access to CONTENTdm Administration with permissions to approve items and build the index of the target collection.
- Data in one of the supported formats (currently CCS Newspapers, CCS Monographs or NDNP Newspapers).
 - Data processed into the METS/ALTO format by the vendors iArchives and Perfect Image have been tested and successfully ingested using the NDNP Newspaper format.
 - Monograph format has been used with ebooks, and has been tested with book metadata from Microsoft Academic Live and CCS.
- Minimum 2 GB RAM.
- Intel Core 2 Duo 2.0 GHz (Intel Core 2 Quad 2.5 GHz recommended).
- 500 GB storage space (fast 100 GB 3.0 Gb/s 7200 RPM disk recommended).

(If you will be running more than one instance of the Flex Loader, we recommend that the storage space is on your local drive.)

- High-speed Ethernet network (gigabit network recommended). A very fast network is recommended if you are going to process and upload large files.

Installing

You need administrative rights on your Windows workstation to install the Flex Loader.

To install the Flex Loader:

1. Double-click the .exe file. Follow the InstallShield screens to install. You will need to enter a valid CONTENTdm license code. Use the same license code as you use for your CONTENTdm Server. (Your code is displayed in CONTENTdm Administration under the Server tab on the About page.)
2. Start the application by going to the Start menu > All Programs > OCLC > CONTENTdm Flex Loader.

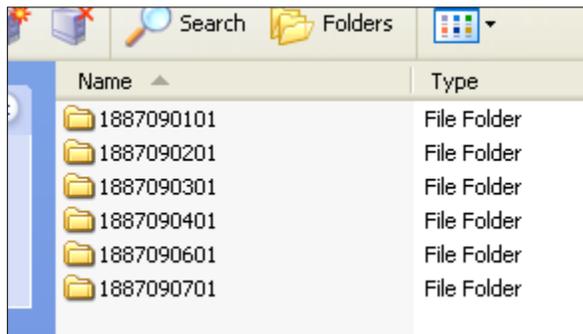
3. The CONTENTdm Flex Loader screen displays.

Setting Up Your Files & Collection

Your data must be in one of the supported formats (currently CCS Newspaper, CCS Monograph or NDNP Newspaper formats).

Before you begin adding files to the Flex Loader, make sure each newspaper or monograph (i.e., all the files related to a single issue or book) is saved in a single folder (subfolders within the single, parent folder can also be used). Files are added to the Flex Loader at the folder level — you cannot add individual files.

For example, the following folder contains six issues of a newspaper. Folder 1887090101 is one issue and is the folder that you would select to add to the processing queue in step two of Using the Flex Loader (below).



The folder 1887090101 contains the following files — these are all of the files for this issue of the newspaper. If you're using subfolders within a single parent folder, one subfolder might contain all the image files and a second subfolder might contain all the XML files.

Name	Size	Type	Date Modified
0125.jp2	5,220 KB	JP2 File	9/25/2008 5:07 PM
0125.pdf	609 KB	Adobe Acrobat Doc...	9/25/2008 5:07 PM
0125.tif	41,724 KB	TIF File	9/25/2008 5:06 PM
0125.xml	894 KB	XML Document	9/25/2008 5:07 PM
0126.jp2	5,165 KB	JP2 File	9/25/2008 5:07 PM
0126.pdf	524 KB	Adobe Acrobat Doc...	9/25/2008 5:07 PM
0126.tif	41,296 KB	TIF File	9/25/2008 5:06 PM
0126.xml	659 KB	XML Document	9/25/2008 5:07 PM
0127.jp2	4,926 KB	JP2 File	9/25/2008 5:07 PM
0127.pdf	406 KB	Adobe Acrobat Doc...	9/25/2008 5:07 PM
0127.tif	39,373 KB	TIF File	9/25/2008 5:06 PM
0127.xml	462 KB	XML Document	9/25/2008 5:07 PM
0128.jp2	5,076 KB	JP2 File	9/25/2008 5:07 PM
0128.pdf	417 KB	Adobe Acrobat Doc...	9/25/2008 5:07 PM
0128.tif	40,570 KB	TIF File	9/25/2008 5:06 PM
0128.xml	406 KB	XML Document	9/25/2008 5:07 PM
0129.jp2	5,082 KB	JP2 File	9/25/2008 5:07 PM
0129.pdf	569 KB	Adobe Acrobat Doc...	9/25/2008 5:07 PM
0129.tif	40,623 KB	TIF File	9/25/2008 5:06 PM
0129.xml	841 KB	XML Document	9/25/2008 5:07 PM
0130.jp2	5,089 KB	JP2 File	9/25/2008 5:07 PM
0130.pdf	358 KB	Adobe Acrobat Doc...	9/25/2008 5:07 PM
0130.tif	40,696 KB	TIF File	9/25/2008 5:07 PM
0130.xml	288 KB	XML Document	9/25/2008 5:07 PM
0131.jp2	5,082 KB	JP2 File	9/25/2008 5:07 PM
0131.pdf	389 KB	Adobe Acrobat Doc...	9/25/2008 5:07 PM
0131.tif	40,625 KB	TIF File	9/25/2008 5:07 PM
0131.xml	412 KB	XML Document	9/25/2008 5:07 PM
0132.jp2	5,122 KB	JP2 File	9/25/2008 5:07 PM
0132.pdf	657 KB	Adobe Acrobat Doc...	9/25/2008 5:07 PM
0132.tif	40,953 KB	TIF File	9/25/2008 5:07 PM
0132.xml	1,038 KB	XML Document	9/25/2008 5:07 PM
1887090101.xml	25 KB	XML Document	9/25/2008 5:06 PM
1887090101_1.xml	70 KB	XML Document	11/21/2008 9:42 PM
articles_1887090101.xml	104 KB	XML Document	3/16/2009 11:53 AM

Please note the following additional file set-up requirements:

- All XML files should have the .xml file extension.
- All ALTO files should contain "ALTO" in the file name.
- For NDNF Newspaper format, the Flex Loader expects two XML files in the METS format: an issue file and an articles file. The primary (issue) file should have a file name that includes an underscore but does NOT start with "articles_" and does NOT contain "-ALTO" (e.g., 1887090101_1.xml). The articles file name should start with "articles_".

Note: If you do not have an articles file, Flex Loader will still process the issue file, and the articles file is treated as empty (the issue is not segmented into articles).

- CCS formats should have "-mets" in the file name of the .xml file.
- All other supported formats should have "mets" (without the hyphen required for CCS formats, see above) in the file name of the .xml file.

- If you want to provide a PDF file of a complete newspaper issue or book, include the PDF file in the folder and name it "all.pdf". The file is uploaded as-is and will become an option for downloading and printing when users view your collection.
- If you want to use Flex Loader to create a PDF file of a complete newspaper issue or book from single-page JP2 or TIFF files (instead of using the all.pdf option described above), check the box labeled **Create Print PDF** in step 2, and then click **Add**. The PDF file will become an option for downloading and printing when users view your collection.

Step 2: Add folders to the processing queue

Use the SHIFT and CTRL keys to select more than one folder

Import from folder

C:\Documents and Settings\dawsone\My Documents\CDM Projects\Fle

..
18950926

Import format

CCS_Newspaper

Create print PDF

Add

Setting Up Your Collection

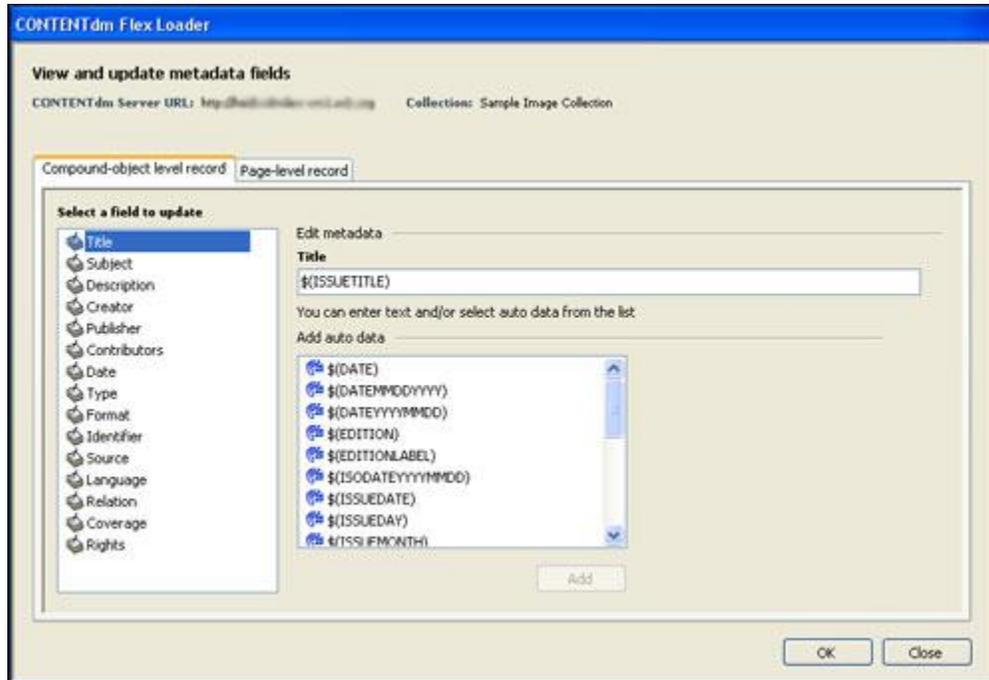
Flex Loader only imports metadata from fields in the XML that are mapped to the collection fields. View and modify the fields, if necessary, in CONTENTdm Administration. We recommend you set up your collection with a full text search field. Although it is not required unless you have transcripts (and want to include the full text in the collection), a full text field enables special features for viewing and searching in your collection (including searching within the document itself). For more information, see **Editing Field Properties**.

Before using the Flex Loader, also confirm your collection configuration settings for the creation of display images. For more information, see the About Image Files section later in this document and the Help topic **Display Image Settings**.

Using the Flex Loader

From the main Flex Loader screen, follow the steps outlined on the screen:

1. Sign in to your CONTENTdm Server and select the collection to which you want to add your files.
2. Optionally, before adding folders to the processing queue, you can view and update metadata fields, including selecting auto data to automatically fill out fields for your items (Flex Loader only imports metadata from fields in the XML that are mapped to the collection fields [mapped field names do not have to match]). Click **View and customize metadata fields for this collection** to get started.

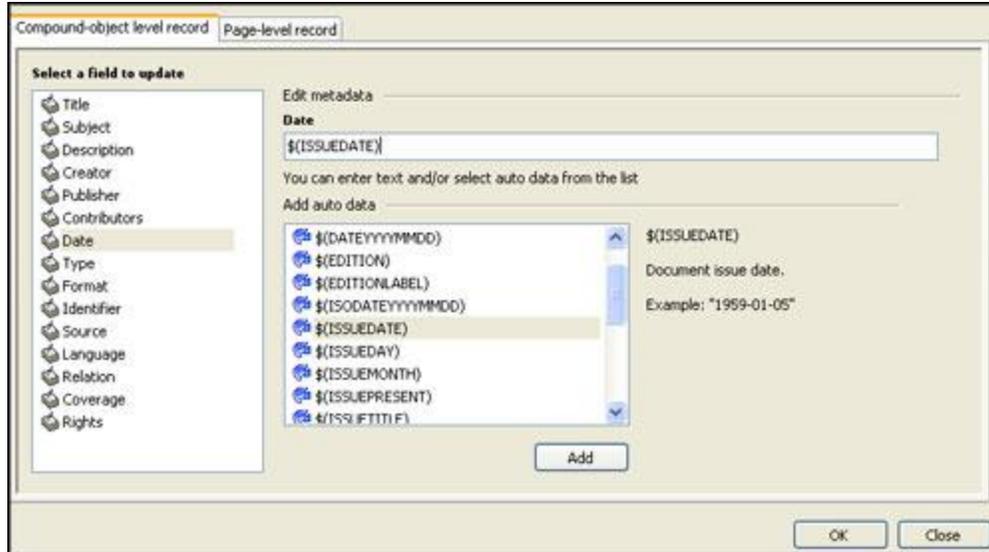


First, select the field to customize. You can enter any combination of text and/or auto data. To add auto data, select the auto data from the list and then click **Add** to add it to the **Edit Metadata** text box for the selected field. When you have finished entering text and/or auto data, click **OK** to save your changes and close the screen. (These metadata field settings will be saved so you can use them for future uploads to the collection.)

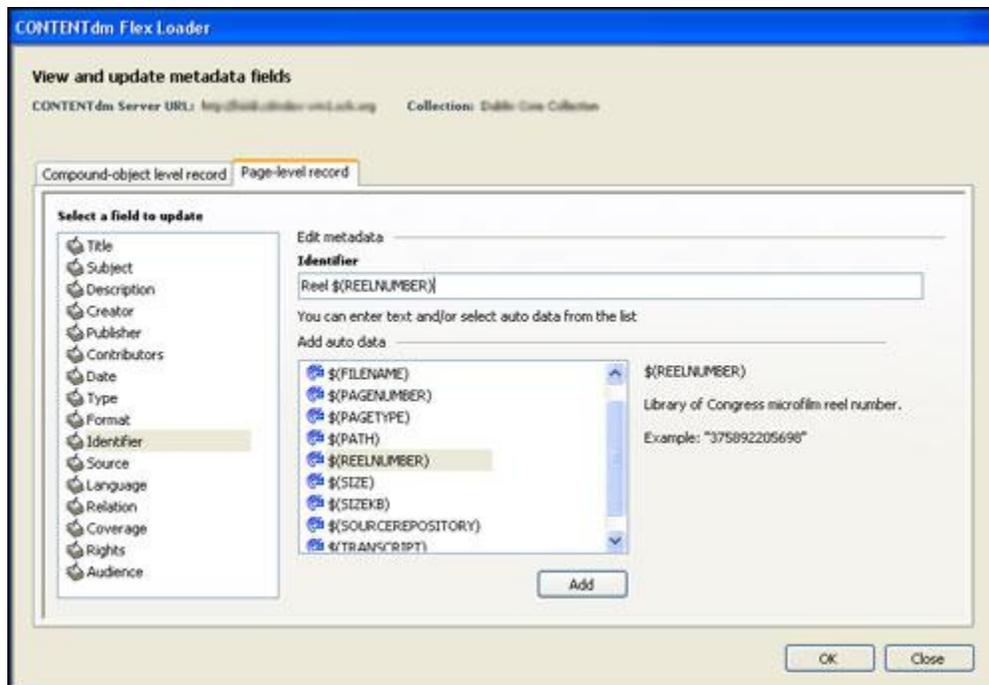
For more information about metadata field mappings from auto data to XML data, see the **Appendix**.

Note: You can edit metadata for the compound object-level record or for the page-level records by selecting the appropriate tab while editing.

For example, you may want to customize the compound-object level record Date field to use the auto data \$(ISSUEDATE).



And you may want to customize the page-level record for the Identifier field to use \$(REELNUMBER) auto data. This example also includes the text label "Reel" preceding the auto data.



Metadata Format Notes:

- Auto data availability may vary depending on format and source files. If auto data is not available for a field, that field is left blank.
- Supported date formats (based on ISO8601 standard) are YYYY-MM-DD, DD.MM.YYYY, MM/DD/YYYY, YYYY-MM and YYYY.
- For NDNF newspaper format: When mapping Title to \$(ISSUETITLE), the LABEL attribute value from the METS XML file is used for the Title field of an issue.

- For CCS monograph and newspaper formats: When mapping Title to \$(ISSUETITLE), the MODS Title value is concatenated and used for the Title field.
3. Then from the main screen, select the folders that contain the files and add them to the processing queue. Remember to specify the format of your files. (For more information about which folder to select, see **Setting Up Your Files.**)
 4. To remove folders from the queue, right-click on the folder name and select **Remove**. A confirmation dialog will confirm your deletion.



To see information about the selected folder, including any error message, click **Information**.



5. Next, click **Start** to process the files. Files are processed by the Web service and added to the approval queue on your CONTENTdm Server.
(Click **Stop** to cancel the upload process. Flex Loader stops the upload after the current folder processing completes.)
6. When processing of all folders has completed, go to CONTENTdm Administration and approve and index the files for the collection.

You can access CONTENTdm Administration by clicking the **Access CONTENTdm Administration** link in the screen to the left of the **Start** button.

For more information, see **Approving Items** and **Building the Text Index**.

Note: You can run more than one instance of the Flex Loader so you can process multiple publications at the same time. We recommend you work from data on a local drive to ensure good performance.

Note: We recommend you plan to work in batches. Processing can take some time, and if you close the Flex Loader before all processing has completed in the queue, the queue list is not retained.

About Image Files

The CONTENTdm Flex Loader uses JPEG2000 or TIFF image file formats.

If JPEG2000 images are available in the import folder, the Flex Loader uses these images first. If JPEG2000 images are not available, the Flex Loader uses TIFF images instead.

When TIFF images are processed, they are converted to display images, based on the collection configuration settings selected on the CONTENTdm Server. (If display image creation is not enabled, display images in JPEG format are automatically created.)

Note: The default configuration on the server is to generate display images in JPEG2000 format, with a compression ratio set to 10:1.

For more information, see [Display Image Settings](#).

Frequently Asked Questions

Q: How do I decide which newspaper format to use?

A: The format you should use depends on the TYPE value in the root of the METS XML file. If the value is "urn:library-of-congress:ndnp:mets:newspaper:issue", you should use the NDNP Newspaper format for processing. The NDNP Newspaper format adheres to the 2009 NDNP Technical [Guidelines](#) and [Specifications](#). If the value is "Newspaper", you should use the CCS Newspaper format.

Q: When I first start the Flex Loader, I get an error message that reads "Web service unavailable" — what do I need to do?

A: This is a temporary situation that should be resolved by OCLC within approximately 15 minutes. Please try starting again later.

Q: What schemas should be followed in formatting my XML files?

A: There are three XML schemas identified in the NDNP specifications that the Flex Loader will process: [METS 1.7](#), [MODS 3.3](#), and [ALTO 1.4](#).

Q: NDNP has many more tagged metadata fields than we would ever use in our collections. Is that a problem? How do you "match up" the CONTENTdm metadata fields with the batch data?

A: You can review the metadata that is currently extracted by clicking the [View and customize metadata fields for this collection](#) link. The auto data name (or snippet) would be replaced. For example, \$(TRANSCRIPT) is replaced by the data extracted from the XML for this field.

Q: Can we use the Flex Loader if we don't meet all of the documented system requirements?

A: The requirements are recommended and represent the tested and supported versions for optimal performance of the Flex Loader. You may be able to run the Flex Loader using a less robust processor, but you may notice significant performance degradation. Additionally, you may be able to run the Flex Loader with a version of Windows other than those recommended in the requirements as the currently tested and supported versions.

Appendix

CONTENTdm Flex Loader Metadata Elements

The following is a list of the data in the XML files that match the auto data fields in the metadata mapping interface in Flex Loader. It also documents which data is automatically extracted and added to CONTENTdm fields, if not mapped using the auto data fields.

1.1. Compound Object Metadata: The following are **not** extracted from XML files:

- \$(DATE) → Current system/computer date, example "Wednesday, June 10, 2009".
- \$(DATEMMDDYY) → Current system/computer date, example "06102009".
- \$(DATEYYYYMMDD) → Current system/computer date, example "20090610".
- \$(ISODATEYYYYMMDD) → Current system/computer date, example "2009-06-10".
- \$(USERNAME) → Username running the application, example "Administrator".
- \$(YEAR) → Current system/computer date, example "2009".
- \$(ZEROAPPEND1) → Appends "0", example, before "1", after "10".
- \$(ZEROAPPEND2) → Appends "00", example, before "1", after "100".
- \$(ZEROAPPEND3) → Appends "000", example, before "1", after "1000".

1.2. Compound Object Metadata: The following **are** extracted from XML files:

- \$(EDITION) → Number of the edition in chronological order, default is "1". Note: In the NDNP specs, this is referred to as "Edition Order."
To locate: mets:mets[@TYPE="urn:library-of-congress:ndnp:mets:newspaper:issue"]/mets:dmdSec[@ID="issueModsbib"]/mets:mdWrap/mets:xmlData/mods:mods/mods:relatedItem/mods:part/mods:detail[@type="edition"]/mods:number
- \$(EDITIONLABEL) → Description of the edition, as printed, example "Final Edition"
To locate: mets:mets[@TYPE="urn:library-of-congress:ndnp:mets:newspaper:issue"]/mets:dmdSec[@ID="issueModsbib"]/mets:mdWrap/mets:xmlData/mods:mods/mods:relatedItem/mods:part/mods:detail[@type="edition"]/mods:caption
- \$(LCCN) → Library of Congress Catalog Number, example "sn83031150"
To locate: mets:mets[@TYPE="urn:library-of-congress:ndnp:mets:newspaper:issue"]/mets:dmdSec[@ID="issueModsbib"]/mets:mdWrap/mets:xmlData/mods:mods/mods:relatedItem/mods:identifier/@type
- \$(ISSUEDATE) → Document issue date, example "1959-01-05"

To locate: Find → mets → dmdSec → MdSecType → MdWrap → XmlData

- \$(ISSUEDAY) → Day of the month that the document was issued, example "05" for the 5th day
To locate: Same as \$(ISSUEDATE) but extracts the day portion of the date, if available.
- \$(ISSUEMONTH) → Month that the document was issued, example "12" for December
To locate: Same as \$(ISSUEDATE) but extracts the month portion of the date, if available.
- \$(ISSUEYEAR) → Year of the document, example "1959"
To locate: Same as \$(ISSUEDATE) but extracts the year portion of the date, if available.
- \$(ISSUEPRESENT) → Valid values are: Present, Not digitized, published; Not digitized, not published; Not digitized, publishing unknown. In the NDNP standard, this is known as "Issue Present Indicator."
To locate: mets:mets[@TYPE="urn:library-of-congress:ndnp:mets:newspaper:issue"]/mets:dmdSec[@ID="issueModsBib"]/mets:mdWrap/mets:xmlData/mods:mods/mods:note
- \$(ISSUETITLE) → Name of the document, example "The Seattle Times."
To locate: Find → mets → extract contents of LABEL attribute.
- \$(TITLEVOLUME) → Document volume, example "5"
To locate: For iArchives: mets:mets[@TYPE="urn:library-of-congress:ndnp:mets:newspaper:issue"]/mets:dmdSec[@ID="issueModsBib"]/mets:mdWrap/mets:xmlData/mods:mods/mods:relatedItem/mods:part/mods:detail[@type="volume"]/mods:number. For CCS: Find → dmdSec → mdWrap → xmlData → mods → relatedItem → identifier → find type="local" → extract contents.
- \$(TITLENUMBER) → Document number, example "2"
To locate: For iArchives: mets:mets[@TYPE="urn:library-of-congress:ndnp:mets:newspaper:issue"]/mets:dmdSec[@ID="issueModsBib"]/mets:xmlData/mods:mods/mods:relatedItem/mods:part/mods:detail[@type="issue"]/mods:number. For CCS: Find → dmdSec → mdWrap → xmlData → mods → titleInfo → partsNumber → extract contents.

2.1. Page Level Metadata: The following are **not** extracted from XML files:

- \$(EXTENSION) → File extension of the pages of the document, example ".jpg".
- \$(FILENAME) → File name without extension, example "clarkcountypage1".
- \$(PAGETYPE) → "Cover" for compound objects or "page" for the pages.
- \$(PATH) → Path to original location of the document.
- \$(SIZE) → Document page size in bytes, example, "95230" bytes.
- \$(SIZEKB) → Document page size in kilobytes, example, "95.230" kilobytes.

2.2. Page Level Metadata: The following **are** extracted from XML files:

- \$(PAGENUMBER) → Page number of document, example "3".
To locate: Find → structMap → div ID="PAGEID" → find TYPE="PAGE" → extract contents of ORDERLABEL attribute.
- \$(REELNUMBER) → Library of Congress microfilm reel number
To locate: mets:mets[@TYPE="urn:library-of-congress:ndnp:mets:newspaper:issue"]/mets:dmdSec[@ID="pageModsBib1"]/mets:mdWrap/mets:xmlData/mods:mods/mods:relatedItem[@type="original"]/mods:identifier[@type="reel number"]
- \$(SOURCEREPOSITORY) → Owner of digitized source; city and state postal abbreviations
To locate: mets:mets[@TYPE="urn:library-of-congress:ndnp:mets:newspaper:issue"]/mets:dmdSec[@ID="pageModsBib1"]/mets:mdWrap/mets:xmlData/mods:mods/mods:relatedItem[@type="original"]/mods:location/mods:physicalLocation/@displayLabel
- \$(TRANSCRIPT) → Page transcript
To locate: Find Layout → PrintSpace → TextBlock → TextLine → String → extract CONTENT attribute.